Prospective Evaluation of Global Earthquake Forecast Models: Two Years of Observations Support Merging Smoothed Seismicity with Geodetic Strain Rates

Anne Strader[1], Maximilian Werner[2], José Bayona[1,3], Philip Maechling[4], Fabio Silva[4], Maria Liukis[5], and Danijel Schorlemmer[1]

[1]GeoForschungsZentrum Potsdam

[2]University of Bristol

[3]Universität Potsdam

[4]University of Southern California

[5]Jet Propulsion Laboratory, NASA

Abstract

The Global Earthquake Activity Rate (GEAR1) seismicity model uses an optimized combination of geodetic strain rates, hypotheses about converting strain rates to seismicity rates from plate tectonics, and earthquake-catalog data to estimate global $m_w \geq 5.767$ shallow (≤ 70 km) seismicity rates. It comprises two parent models: a strain rate-based model, and a smoothed-seismicity based model. The GEAR1 model was retrospectively evaluated and calibrated using earthquake data from 2005–2012, resulting in a preferred log-linear, multiplicative combination of the parent forecasts. Since October 1, 2015, the GEAR1 model has undergone prospective evaluation within the Collaboratory for the Study of Earthquake Predictability (CSEP) testing center.

We present initial prospective forecast test results for the GEAR1 model, its tectonic and seismicity components, and for the first iteration of the strain rate-based model, during the period October 1, 2015 to September 7, 2017. Observed earthquakes are consistent with the GEAR1 forecast, supporting a near-even contribution from geodetic strain rates and smoothed seismicity in constraining long-term global shallow earthquake rates. Comparative test results likewise support

that GEAR1 is more informative than either of its components alone. The tectonic forecasts do not effectively anticipate observed spatial earthquake distribution, largely due to intraplate earthquakes and a higher-than-implied degree of spatial earthquake clustering.

Introduction

Earthquake-source models that reliably forecast long-term seismicity rates are imperative to seismic risk mitigation. As they form the foundation of seismic hazard models (Anderson and Biasi, 2016; Cornell, 1968), the development and evaluation of seismicity models is essential to improving ground shaking estimates. Such estimates are directly used worldwide as an input for building codes, including nuclear facilities, and for urban planning. Accurate and reliable forecasting of long-term seismicity rates allows societies to identify regions at risk of catastrophic earthquake damage and effectively invest in building and infrastructure safety. Global seismicity models are advantageous in this respect because sufficient earthquakes occur to rank global forecasts with a 5.8–7.0 magnitude threshold after only 1-8 years of testing (Bird et al., 2015).

Until recently, global seismicity models used as input for Probabilistic Seismic Hazard Analysis (PSHA) were primarily based on a combination of earthquake-catalog and fault data (Giardini, 2014; Giardini et al., 1999). These seismicity models extrapolated previous seismicity to the future, and assumed that all faults that could potentially yield large, catastrophic earthquakes were known. In recent years, the quality and quantity of available geodetic data have increased considerably and are beginning to be incorporated into earthquake source models such as the Uniform California Earthquake Rupture forecast (UCERF2) (Field et al., 2009) and UCERF3 (Field et al., 2015). These data can provide information about tectonic deformation that cannot be supplied by instrumental or historical earthquake catalogs, particularly in regions where only few seismicity data are available (Kreemer et al., 2014; Bird et al., 2015; Bird and Kreemer, 2015).

Seismicity Models

To use geodetic data in forecasting time-invariant global seismicity rates, Bird and Kreemer (2015) developed the SHIFT_GSRM2f model. SHIFT_GSRM2f is based on the second iteration of the Global Strain Rate Map (GSRM2.1), a global model of continuous strain rates constrained by 22,415 interseismic GPS velocities (Kreemer et al., 2014). Using the Seismic Hazard Inferred from Tectonics (SHIFT) hypotheses introduced by Bird and Liu (2007), Bird and Kreemer converted geodetic strain rates to long-term seismicity rates by multiplying each strain rate tensor by the elastic shear modulus, spatial bin area, geometric factor, and coupled seismogenic thickness. They produced six tectonics-based seismicity models of varying complexity (SHIFT_GSRM2a-f). SHIFT_GSRM2f, the most complex of the resulting seismicity models, recognizes five tectonic zones, applies asymmetric Gaussian smoothing to subduction zone strain rates and symmetric smoothing to all other offshore plate boundaries, and accounts for velocity-dependent coupled seismogenic thickness.

The Global Earthquake Activity Rate Model (GEAR1) is a hybrid seismicity model that uses a combination of geodetic strain rates and earthquake-catalog data to forecast long-term seismicity (Bird et al., 2015). It is a multiplicative log-linear combination of the SHIFT_GSRM2f and Kagan-Jackson Smoothed Seismicity (KJSS) models (Kagan and Jackson, 1994, 2000, 2011). The KJSS model was developed by smoothing shallow ($\leq$ 70 km) $m_w \geq 5.767$ earthquakes during the period 1977–2004 from the Global Centroid Moment Tensor (CMT) earthquake catalog (Ekström et al., 2012). Although strain rates capture seismicity in regions where earthquake-catalog data are unavailable, earthquake catalogs provide information about intraplate seismicity, where strain rates are assumed to be zero according to the GSRM2.1.

Using global CMT seismicity from 2005–2012, Bird et al. (2015) retrospectively evaluated various combinations of the SHIFT_GSRM2f and KJSS models. They tested three main combinations of tectonics and seismicity: a weighted linear combination, a multiplicative log-linear combination, and selection in each spatial bin of the model with the greater forecasted seismicity rate. They optimized the GEAR1 model by maximizing the I1 ("success") score, or mean information gain per earthquake of GEAR1 over a uniform Poisson model (Kagan, 2009). Testing multiple weightings of tectonics and smoothed seismicity, they determined that the optimal GEAR1 model was a multiplicative log-linear combination of SHIFT_GSRM2f and KJSS, with a weight of 0.6 (d=0.6 in the following equation) assigned to smoothed seismicity:

$$H_{ij} = N\{\sup\left[\left(S_{ij}^{d} \cdot T_{ij}^{1-d}\right), f\right]\} \tag{1}$$

where H_{ij} is the total forecasted earthquake rate in the spatial bin with latitude i and longitude j. S_{ij} is the rate forecasted by KJSS and T_{ij} is the rate forecasted by SHIFT_GSRM2f. f is a baseline seismicity rate defined as $\min\left[\min\left(S_{ij}\right), \min\left(T_{ij}\right)\right]$, in order to account for earthquakes in unexpected locations. $N\{\}$ is an operator that normalizes the total earthquake rate to equal the total rate observed during 1977–2003 according to the CMT catalog, so that the models project the mean global seismicity rate into the future. Figure 1 displays maps of earthquake rates calculated from all three global models, as well as SHIFT_GSRM (Bird et al., 2010), the predecessor of SHIFT_GSRM2F.

Seismicity Forecast Evaluation

Although successful retrospective test results are necessary to detect initial seismicity model limitations, such tests are prone to bias because data included in calibrating the model are also used to test it. Only truly prospective evaluation is considered rigorous, because it has the ability to estimate the forecasting power of a seismicity model (Schorlemmer et al., 2007; Zechar et al., 2010b; Strader et al., 2017). Since October 1, 2015, seismicity forecasts produced by GEAR1, SHIFT_GSRM2f, and KJSS have undergone prospective testing within the Collaboratory for the Study of Earthquake Predictability (CSEP). We also evaluate the SHIFT_GSRM model, based on the first version of the GSRM (Kreemer et al., 2003). GSRM2.1 improves upon the original GSRM through its incorporation of additional geodetic data, systematic data processing, additional modeled plates and plate boundaries, and a finer spatial grid (Bird and Kreemer, 2015; Kreemer et al., 2014).

Recent Results

In this study, we present prospective testing results from October 1, 2015 to September 7, 2017. We evaluate the consistency of each forecast's total number of earthquakes, spatial earthquake distribution and magnitude distribution against earthquakes observed during the evaluation period. To determine each GEAR1 component's contribution to constraining seismicity rates and prospectively assess the stability of GEAR1's optimization, we compare GEAR1's forecast performance with those of SHIFT_GSRM2f and KJSS. We also compare the performance of SHIFT_GSRM and SHIFT_GSRM2f to determine if the additional geodetic data used in developing SHIFT_GSRM2f significantly improves forecast performance. Using these CSEP test results, we identify primary model strengths and weaknesses, and suggest possible model improvements in future iterations of GEAR1 and SHIFT_GSRM2f.

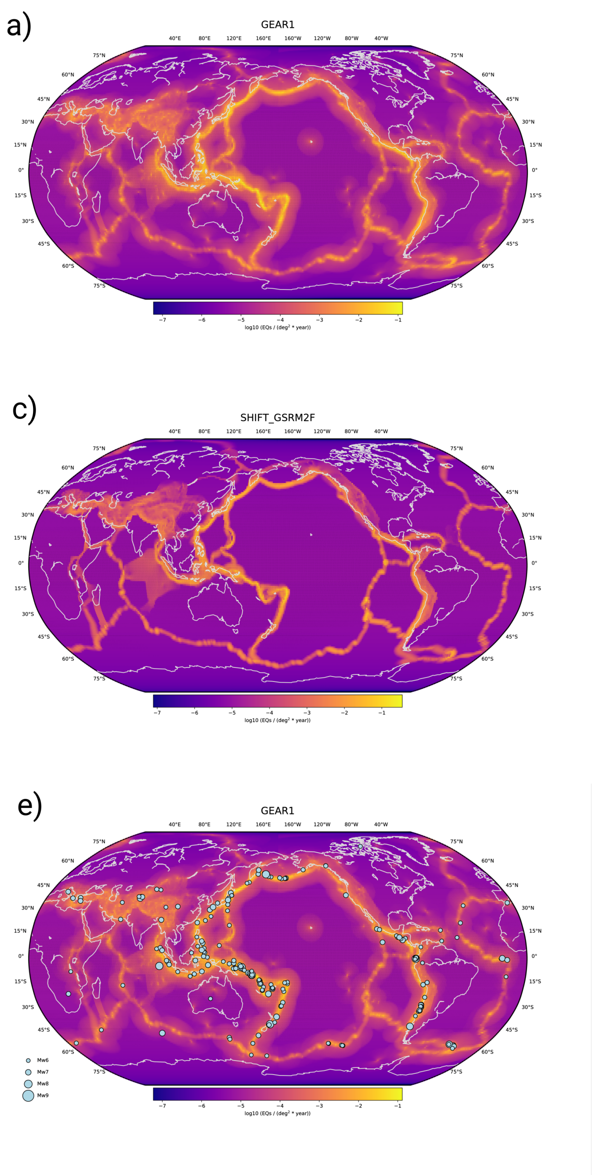

Figure 1: Forecast maps showing $m_w \geq 5.767$ earthquake epicentroid rates (log10(eqs/year) in each $0.1° \times 0.1°$ spatial bin): a) the preferred GEAR1 model (Bird et al., 2015), b) SHIFT_GSRM (Bird et al., 2010), c) SHIFT_GSRM2f (Bird and Kreemer, 2015), d) KJSS (Kagan and Jackson, 2011). We determine epicentroid rate densities from a loglinear, multiplicative combination of the SHIFT_GSRM2F and KJSS parent forecasts, with exponent d = 0.6. Bright yellow and orange areas indicate regions with elevated seismicity rates. Map e) displays epicenters of observed earthquakes during the evaluation period (overlying the GEAR1 forecast map).

Data

209 $m_w \geq 5.767$ earthquakes (0–70 km depth) occurred during the evaluation period, approximately half the number of earthquakes observed per two-year period during the 2005–2012 GEAR1 calibration period. We used the undeclustered CMT catalog, consistent with earthquakes used to generate and calibrate GEAR1 (Bird et al., 2015). Notable earthquakes include the April 16, 2016 m_w 7.8 Ecuador earthquake, November 14, 2016 m_w 7.8 Kaikoura earthquake in New Zealand, and December 17, 2016 m_w 7.9 Papua New Guinea earthquake. The earthquake catalog is included in Table S1 (available in the electronic supplement to this article).

Methods

CSEP Testing Center Framework

Each CSEP testing center (Schorlemmer and Gerstenberger, 2007) allows for earthquake forecasts generated from seismicity models to be evaluated in a specified region during a specified evaluation period. In a forecasting experiment, the testing region is divided into spatiomagnitude bins based on longitude, latitude and magnitude increments. Seismicity modellers submit forecasts to CSEP in the form of expected numbers of earthquakes (assuming a Poisson seismicity distribution in each spatial bin) during the evaluation period in each spatiomagnitude bin. Forecasting experiments are conducted in a transparent, reproducible environment, with earthquake data used to test forecasts provided by an authoritative and independent source (Schorlemmer and Gerstenberger, 2007; Zechar et al., 2010b). A brief overview of the consistency and comparative tests (with corresponding literature) used to evaluate earthquake forecasts is provided below.

Comparative Tests

One evaluates the relative performance of forecast pairs by measuring the rate-corrected information gain per earthquake of one forecast over another (Rhoades et al., 2011). We apply the Student's paired T-test, defining the null hypothesis that two forecasts do not perform significantly differently, and the alternate hypothesis that one forecast significantly outperforms the other. If the mean information gain significantly differs from the scaled difference in forecasted earthquake numbers between two forecasts, one forecast is significantly more informative than the other.

In the case that the information gain scores at observed earthquake locations are not normally distributed, the non-parametric W-test evaluates the median information gain per earthquake rather than the mean. This test only requires that the information gain distribution is symmetric, and increases in power with increasing numbers of observed earthquakes (Rhoades et al., 2011).

Consistency Tests

Using a suite of likelihood consistency tests, CSEP evaluates the consistency of forecasted and observed seismicity during the experiment's evaluation period. Tests are based on the likelihood of observed seismicity patterns, given forecasted seismicity rates.

A forecast's log-likelihood score is a metric, based on the Poisson distribution, used to evaluate the consistency of forecasted seismicity patterns with observed earthquakes (Schorlemmer et al., 2007). Greater log-likelihood scores indicate greater consistency, corresponding to a higher probability of the forecast generating a seismicity distribution similar to observations (indicating greater ability to forecast earthquakes).

Consistency with observed seismicity, or the log-likelihood score, can be decomposed into three dimensions: number of earthquakes, magnitude and spatial distributions. The tests of consistency for these dimensions (N-test, M-test and S-test, respectively) are derived directly from the likelihood, or L-test (Zechar et al., 2010a). Because the number of earthquakes has the greatest impact on a forecast's log-likelihood score, the conditional likelihood (CL) test provides information about a forecast's spatial and magnitude distribution with information regarding the total number of earthquakes removed (Werner et al., 2011).

Results

Model Comparison and Ranking

GEAR1 significantly outperforms both of its individual components, according to the t- and W-test results (Figure 2). This is based on the information gain at a selected significance level of 0.05. The forecast's superior performance to its individual components during the prospective evaluation supports the near-even weighting of SHIFT_GSRM2f and KJSS, as well as the selected log-linear combination of the models. Given that the multiplicative model captures two independent factors necessary for earthquake triggering — continuous lithospheric deformation as well as sudden static or dynamic stress changes (Bird et al., 2015)—the stability of the optimized GEAR1 model is promising. Due to the increased number of earthquakes included in global forecast evaluation, the test results are a more conclusive indicator of relative forecast performance compared to regional forecasting experiments such as RELM.

SHIFT_GSRM displays a small mean information gain over SHIFT_GSRM2f; however, this difference is not statistically significant.

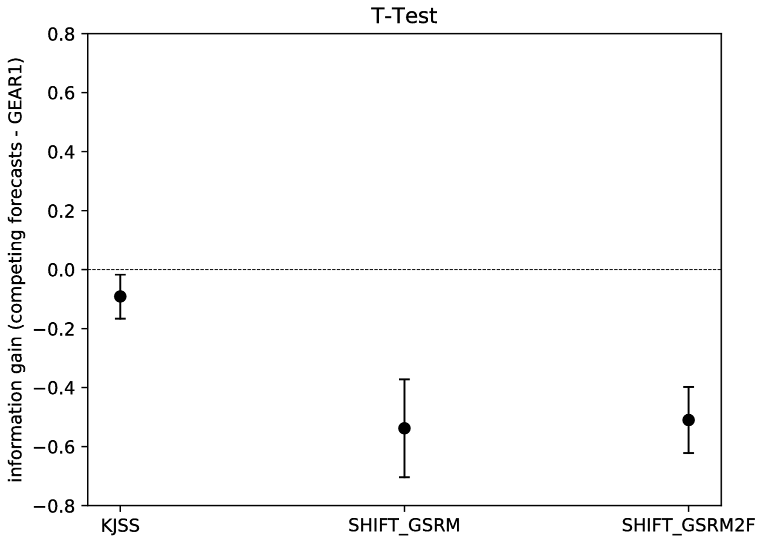

Figure 2: T- and W-test results comparing GEAR1's performance with those of KJSS, SHIFT_GSRM and SHIFT_GSRM2f. GEAR1 significantly outperforms both tectonics forecasts as well as KJSS, supporting a near-even combination of geodetic and earthquake-catalog data to best constrain short- to mid-term seismicity rates. Circles correspond to observed mean information gain per earthquake, while vertical lines display the range of 0.05 significance level mean information gain values; the ranges for all tectonic and seismicity component forecasts are all below zero. The mean information gain per earthquake is displayed on the y-axis; the dashed horizontal line corresponds to the scaled mean information gain (that is, no difference in performance between forecast pairs). Although not displayed, the W-test results corroborate all T-test results.

Consistency Tests

The observed total earthquake number is consistent with all forecasts, according to the N-test results (Figure 3). Because all models had similar spatial seismicity rate integrals, it is unsurprising that the models passed the N-test simultaneously. All models were normalized to forecast near-equal total earthquake rates so that annual fluctuations in seismicity would not favor one forecast over another (Bird et al., 2015).

Likewise, the observed earthquake magnitude distribution is consistent with those of all forecasts, as indicated by the M-test results. Prior to combining the KJSS and SHIFT_GSRM2f models, Bird et al. (2015) scaled the KJSS model to have an equal seismicity rate to the SHIFT_GSRM2f model. Both models' magnitude-frequency distributions comprised unions of various tapered Gutenberg-

Richter distributions based on the SHIFT philosophy (Bird and Kagan, 2004). Therefore, the four models have similar log-likelihood (M) scores.

Both GEAR1 and KJSS forecast spatial earthquake distributions consistent with observations, as indicated by the S-test; however, both tectonic models do not. One reason for the difference in performance is that the tectonic models do not adequately capture intraplate seismicity, as indicated by differences in log-likelihood scores (see Table S1, available in the electronic supplement to this article) in spatiomagnitude bins containing these earthquakes. Additionally, forecasted seismicity in SHIFT_GSRM2F is highly concentrated along plate boundaries compared to KJSS, as shown in concentration plots comparing cumulative observed and forecasted seismicity for both forecasts (Figure 4).

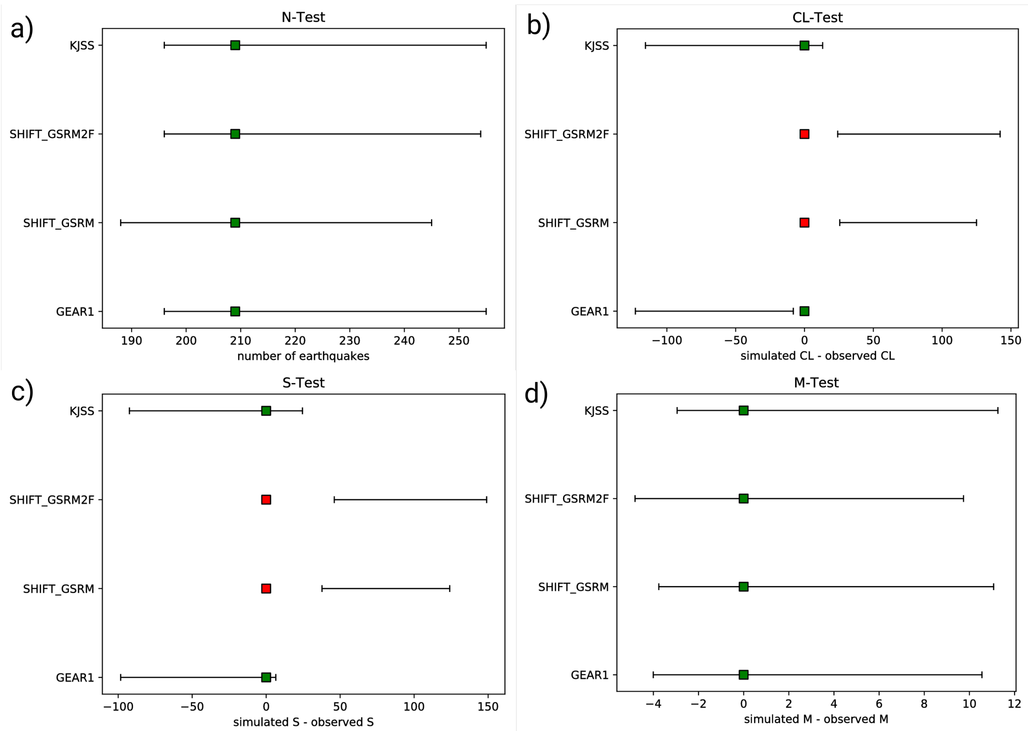

Figure 3: Consistency test results for the GEAR1, SHIFT_GSRM, SHIFT_GSRM2F, and KJSS forecasts during the period October 1, 2015–September 7, 2017: a) N-test, b) CL-test, c) S-test, d) M-test. Green squares indicate that observations are consistent with the forecasts, while red squares indicate significant inconsistencies between forecasted and observed seismicity. All forecasts pass the N- and M-tests; however, SHIFT_GSRM and SHIFT_GSRM2F forecast spatial earthquake distributions inconsistent with observed seismicity, thus failing the CL- and S-tests. For the N-test, the ordinate displays the number of earthquakes; for all other tests, the log-likelihood score is displayed. The squares indicate number of observed earthquakes (in the case of the N-test) or the log-likelihood score calculated using observed earthquakes (all other tests). The horizontal lines indicate the range of observed earthquake numbers or log-likelihood scores within the 0.05 significance level.

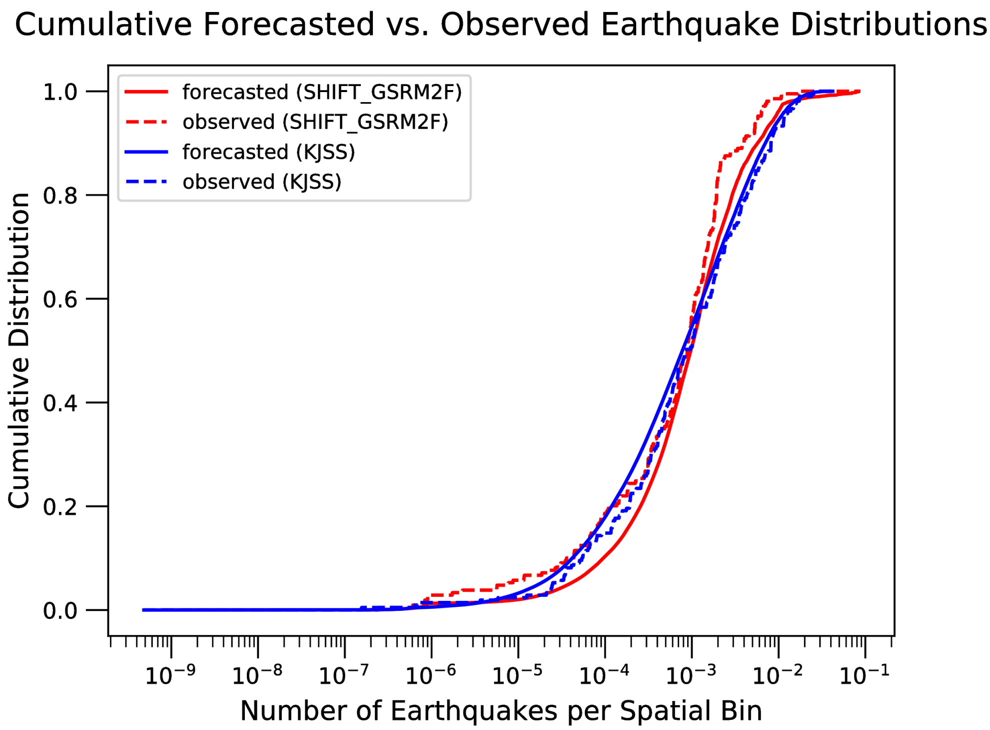

Figure 4: Concentration plots displaying the cumulative forecasted versus observed seismicity distribution for SHIFT_GSRM2F (red curves) and KJSS (blue curves). The observed seismicity distribution for SHIFT_GSRM2F is shifted slightly to the left of the forecasted distribution, indicating that the forecasted seismicity is too localized. For KJSS, the two distributions are in closer agreement, though the forecasted seismicity distribution is slightly over-smoothed.

Discussion and Conclusions

We prospectively evaluated forecasts generated from the GEAR1 model, its components SHIFT_GSRM2f and KJSS, and the original SHIFT_GSRM model from October 1, 2015 to September 7, 2017 within the CSEP testing center. The total earthquake number, spatial and magnitude distributions forecasted by GEAR1 were all consistent with observed seismicity. Additionally, GEAR1 outperformed component models SHIFT_GSRM2f and KJSS, supporting the inclusion both geodetic strain rates and earthquake-catalog data to better constrain seismicity rates.

Despite the incorporation of numerous geodetic data and additional tectonic plates and plate boundaries, SHIFT_GSRM2f did not significantly outperform SHIFT_GSRM. Although SHIFT_GSRM2f incorporates more geodetic velocities and provides a higher-resolution spatial strain-rate distribution than SHIFT_GSRM, it did not account for intraplate seismicity or imply a substantially higher degree of earthquake clustering. Both tectonic models forecasted spatial seismicity patterns inconsistent with observed earthquakes, indicating localized over/underprediction.

GEAR1's superior performance to the SHIFT_GSRM and SHIFT_GSRM2f models indicates that geodetic data supplements, but does not replace earthquake-catalog data. This is corroborated by the rejection of KJSS in favor of GEAR1, which supports the combination and weighting of tectonics and smoothed seismicity inferred from 2005–2012 seismicity when calibrating GEAR1. Because of the increased power of the global tests relative to regional forecasting experiments such as RELM (Strader et al., 2017), such evaluations may not require decades to build confidence. Future prospective testing will more clearly elucidate how global forecast performance stability is affected by the duration of the evaluation period, in particular the effects of earthquake clustering from KJSS.

In addition to the tectonic forecasts' limitations in forecasting intraplate seismicity patterns, Bird et al. (2015) indicate that the SHIFT_GSRM and SHIFT_GSRM2f models tend to underpredict seismicity within subduction zones prior to applying empirical calibration factors to account for the effect of the dip angle on the geometric factor. Another possible factor that could have contributed to spatial seismicity distribution discrepancies is the increase in seismicity rates within subduction zones from 2004 to 2016 (including GEAR1's calibration period and part of the evaluation period) compared to 1977–2003. As temporal fluctuations in seismicity within specific regions or tectonic zones result in temporarily-lowered log-likelihood scores during short forecast-evaluation periods, further prospective testing is necessary to determine whether observed spatial inconsistencies are significant over a time period relevant for seismic hazard analysis.

Inconsistencies in spatial seismicity patterns between the SHIFT_GSRM or SHIFT_GSRM2f model and observed seismicity may also be caused by inaccurate or low-resolution physical input parameters used to convert strain rates to seismic moment. Bayona et al. (in prep) are currently investigating how varying input parameters in the SHIFT_GSRM2f model (for example: subduction dip angle, coupled seismogenic thickness, and corner magnitude) affects ratios of forecasted to observed earthquakes in subduction zones. Further testing of individual regions and the subsequent development and prospective evaluation of updated SHIFT_GSRM and GEAR models will indicate the extent to which forecast performance is sensitive to parameters used to convert strain rates to earthquake rates.

Initial prospective evaluation of the GEAR1 model, its individual components, and the SHIFT_GSRM model supports the incorporation of both geodetic strain rates and earthquake-catalog data in earthquake source models used for input in PSHA. The global forecast ranking indicates that geodetic strain rates provide information about long-term seismicity rates missing from available instrumental earthquake catalogs. Conversely, earthquake-catalog data better constrain intraplate seismicity and capture clustering patterns along plate boundaries. These findings are similar to those from regional forecasting experiments such as the RELM experiment (Strader et al., 2017). Seismicity models such as UCERF2 that incorporate geodetic strain rates not only forecast seismicity patterns consistent with observations, but their forecast performance over consecutive five-year time intervals is more consistent than for models solely based on smoothed seismicity (Helmstetter et al., 2007). In the future, the predictive skill of SHIFT_GSRM2f could be improved further with more region-specific calibration of parameters used to convert strain rates to seismicity rates, as well as further investigation of strain rate variations in intraplate regions. However, prospective global forecast evaluation results support incorporation of geodetic strain rates in constraining long-term seismicity rates.

Data and Resources

The observation earthquake catalog was obtained from the global CMT earthquake catalog by the CSEP testing center. All forecast evaluation tests except for the concentration plots were conducted using miniCSEP, available as open-source software from CSEP.

Acknowledgments

The authors thank Peter Bird, David D. Jackson, Corné Kreemer, Yan Kagan and Ross Stein for the development of the GEAR1 model and its components, as well as valuable discussion on interpreting prospective forecast test results. Additionally, Peter Bird developed and provided code to generate the GEAR1 forecast, which he adapted to generate the KJSS smoothed seismicity forecast. We also thank Thomas Beutin and John Yu for technical support with running the CSEP tests. This study was made possible by the open-source community and the Linux operating system.

This research was funded by the Global Earthquake Model (GEM) foundation and King Abdullah University of Science and Technology (KAUST) research grant URF/1/2160-01-01.

References

Anderson, J. G. and Biasi, G. P. What is the Basic Assumption for Probabilistic Seismic Hazard Assessment? *Seismological Research Letters*, 87(2A):323–326, 2016.

Bird, P. and Kagan, Y. Y. Plate-Tectonic Analysis of Shallow Seismicity: Apparent Boundary Width, Beta, Corner Magnitude, Coupled Lithosphere Thickness, and Coupling in Seven Tectonic Settings. *Bulletin of the Seismological Society of America*, 94(6):2380–2399, 2004.

Bird, P. and Kreemer, C. Revised Tectonic Forecast of Global Shallow Seismicity Based on Version 2.1 of the Global Strain Rate Map. *Bulletin of the Seismological Society of America*, 105(1): 152–166, 2015.

Bird, P. and Liu, Z. Seismic Hazard Inferred from Tectonics: California. *Seismological Research Letters*, 78(1):37–48, 2007.

Bird, P., Kreemer, C., and Holt, W. E. A Long-Term Forecast of Shallow Seismicity Based on the Global Strain Rate Map. *Seismological Research Letters*, 81(2):184–194, 2010.

Bird, P., Jackson, D. D., Kagan, Y. Y., Kreemer, C., and Stein, R. S. GEAR1: A Global Earthquake Activity Rate Model Constructed from Geodetic Strain Rates and Smoothed Seismicity. *Bulletin of the Seismological Society of America*, 105(5):2538–2554, 2015.

Cornell, C. A. Engineering Seismic Risk Analysis. *Bulletin of the Seismological Society of America*, 58(5):1583–1606, 1968.

Ekström, G., Nettles, M., and Dziewoński, A. M. The Global CMT Project 2004-2010: Centroid Moment Tensors for 13,017 Earthquakes. *Physics of the Earth and Planetary Interiors*, 200–201: 1–9, 2012.

Field, E. H., Dawson, T. E., Felzer, K. R., Frankel, A. D., Gupta, V., Jordan, T. H., Parsons, T., Petersen, M. D., Stein, R. S., Weldon II, R. J., and Wills, C. J. Uniform California Earthquake Rupture Forecast, Version 2 (UCERF2). *Bulletin of the Seismological Society of America*, 99(4): 2053–2107, 2009.

Field, E. H., Biasi, G. P., Bird, P., Dawson, T. E., Felzer, K. R., Jackson, D. D., Johnson, K. M., Jordan, T. H., Madden, C., Michael, A. J., Milner, K. R., Page, M. T., Parsons, T., Powers, P. M., Shaw, B. E., Thatcher, W. R., Weldon II, R. J., and Zeng, Y. Long-Term Time-Dependent Probabilities for the Third Uniform California Earthquake Rupture Forecast (UCERF3). *Bulletin of the Seismological Society of America*, 105(2A):511–543, 2015.

Giardini, D. Mapping Europe's Seismic Hazard. *EOS*, 95(29):261–268, 2014.

Giardini, D., Grünthal, G., Shedlock, K. M., and Zhang, P. The GSHAP Global Seismic Hazard Map. *Annals of Geophysics*, 42(6):1225–1230, 1999.

Helmstetter, A., Kagan, Y. Y., and Jackson, D. D. High-Resolution Time-Independent Grid-Based Forecast for $M \geq 5$ Earthquakes in California. *Seismological Research Letters*, 78(1):78–86, 2007.

Kagan, Y. Y. Testing Long-Term Earthquake Forecasts: Likelihood Methods and Error Diagrams. *Geophysical Journal International*, 177(2):532–542, 2009.

Kagan, Y. Y. and Jackson, D. D. Long-Term Probabilistic Forecasting of Earthquakes. *Journal of Geophysical Research: Solid Earth*, 99(B7):13685–13700, 1994.

Kagan, Y. Y. and Jackson, D. D. Probabilistic Forecasting of Earthquakes. *Geophysical Journal International*, 143(2):438–453, 2000.

Kagan, Y. Y. and Jackson, D. D. Global Earthquake Forecasts. *Geophysical Journal International*, 184(2):759–776, 2011.

Kreemer, C., Holt, W. E., and Haines, A. J. An Integrated Global Model of Present-Day Plate Motions and Plate Boundary Deformation. *Geophysical Journal International*, 154(1):8–34, 2003.

Kreemer, C., Blewitt, G., and Klein, E. C. A Geodetic Plate Motion and Global Strain Rate Model. *Geochemistry, Geophysics, Geosystems*, 15(10):3849–3889, 2014.

Rhoades, D. A., Schorlemmer, D., Gerstenberger, M. C., Christophersen, A., Zechar, J. D., and Imoto, M. Efficient Testing of Earthquake Forecasting Models. *Acta Geophysica*, 59(4):728–747, 2011.

Schorlemmer, d. and Gerstenberger, M. RELM Testing Center. *Seismological Research Letters*, 78 (1):30–36, 2007.

Schorlemmer, D., Gerstenberger, M. C., Wiemer, S., Jackson, D. D., and Rhoades, D. A. Earthquake Likelihood Model Testing. *Seismological Research Letters*, 78(1):17–29, 2007.

Strader, A. E., Schneider, M., and Schorlemmer, D. Prospective and Retrospective Evaluation of Five-Year Earthquake Forecast Models for California. *Geophysical Journal International*, 211 (1):239–251, 2017.

Werner, M. J., Helmstetter, A., Jackson, D. D., and Kagan, Y. Y. High-Resolution Long-Term and Short-Term Earthquake Forecasts for California. *Bulletin of the Seismological Society of America*, 101(4):1630–1648, 2011.

Zechar, J. D., Gerstenberger, M. C., and Rhoades, D. A. Likelihood-Based Tests for Evaluating Space-Rate-Magnitude Earthquake Forecasts. *Bulletin of the Seismological Society of America*, 100(3):1184–1195, 2010a.

Zechar, J. D., Schorlemmer, D., Liukis, M., Yu, J., Euchner, F., Maechling, P. J., and Jordan, T. H. The Collaboratory for the Study of Earthquake Predictability Perspective on Computational Earthquake Science. *Concurrency and Computation Practice and Experience*, 22(12):1836–1847, 2010b.